

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: GIJZEN, Mark
- (ii) TITLE OF INVENTION: SEED COAT SPECIFIC DNA REGULATORY REGION  
AND PEROXIDASE
- (iii) NUMBER OF SEQUENCES: 2
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: NIXON & VANDERHYE P.C.
  - (B) STREET: 8th Floor, 1100 North Glebe Road
  - (C) CITY: Arlington
  - (D) STATE: Virginia
  - (E) COUNTRY: United States
  - (F) ZIP: 22201-4714
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER:
  - (B) FILING DATE: 26-SEP-1997
  - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 08/723,414
  - (B) FILING DATE: 30-SEP-1996
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: BYRNE, Thomas E.
  - (B) REGISTRATION NUMBER: 32,205
  - (C) REFERENCE/DOCKET NUMBER: 76-105
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: (703) 816-4021
  - (B) TELEFAX: (703) 816-4100

(2) INFORMATION FOR SEQ ID NO: 1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1244 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION:1..1056

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide

(B) LOCATION: 1..77

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

ATG GGT TCC ATG CGT CTA TTA GTA GTG GCA TTG TTG TGT GCA TTT GCT 48  
Met Gly Ser Met Arg Leu Leu Val Val Ala Leu Leu Cys Ala Phe Ala  
1 5 10 15

ATG CAT GCA GGT TTT TCA GTC TCT TAT GCT CAG CTT ACT CCT ACG TTC 96  
Met His Ala Gly Phe Ser Val Ser Tyr Ala Gln Leu Thr Pro Thr Phe  
20 25 30

TAC AGA GAA ACA TGT CCA AAT CTG TTC CCT ATT GTG TTT GGA GTA ATC 144

Tyr Arg Glu Thr Cys Pro Asn Leu Phe Pro Ile Val Phe Gly Val Ile

35 40 45

TTC GAT GCT TCT TTC ACC GAT CCC CGA ATC GGG GCC AGT CTC ATG AGG 192

Phe Asp Ala Ser Phe Thr Asp Pro Arg Ile Gly Ala Ser Leu Met Arg

50

55

60

CTT CAT TTT CAT GAT TGC TTT GTT CAA GGT TGT GAT GGA TCA GTT TTG

240

Leu His Phe His Asp Cys Phe Val Gln Gly Cys Asp Gly Ser Val Leu

65

70

75

80

CTG AAC AAC ACT GAT ACA ATA GAA AGC GAG CAA GAT GCA CTT CCA AAT

288

Leu Asn Asn Thr Asp Thr Ile Glu Ser Glu Gln Asp Ala Leu Pro Asn

85

90

95

ATC AAC TCA ATA AGA GGA TTG GAC GTT GTC AAT GAC ATC AAG ACA GCG

336

Ile Asn Ser Ile Arg Gly Leu Asp Val Val Asn Asp Ile Lys Thr Ala

100

105

110

GTG GAA AAT AGT TGT CCA GAC ACA GTT TCT TGT GCT GAT ATT CTT GCT

384

Val Glu Asn Ser Cys Pro Asp Thr Val Ser Cys Ala Asp Ile Leu Ala

115

120

125

ATT GCA GCT GAA ATA GCT TCT GTT CTG GGA GGA GGT CCA GGA TGG CCA

432

Ile Ala Ala Glu Ile Ala Ser Val Leu Gly Gly Gly Pro Gly Trp Pro

130

135

140

GTT CCA TTA GGA AGA AGG GAC AGC TTA ACA GCA AAC CGA ACC CTT GCA

480

Val Pro Leu Gly Arg Arg Asp Ser Leu Thr Ala Asn Arg Thr Leu Ala

145

150

155

160

AAT CAA AAC CTT CCA GCA CCT TTC TTC AAC CTC ACT CAA CTT AAA GCT

528

Asn Gln Asn Leu Pro Ala Pro Phe Phe Asn Leu Thr Gln Leu Lys Ala

165

170

175

TCC TTT GCT GTT CAA GGT CTC AAC ACC CTT GAT TTA GTT ACA CTC TCA	576
Ser Phe Ala Val Gln Gly Leu Asn Thr Leu Asp Leu Val Thr Leu Ser	
180 185 190	
GGT GGT CAT ACG TTT GGA AGA GCT CGG TGC AGT ACA TTC ATA AAC CGA	624
Gly Gly His Thr Phe Gly Arg Ala Arg Cys Ser Thr Phe Ile Asn Arg	
195 200 205	
TTA TAC AAC TTC AGC AAC ACT GGA AAC CCT GAT CCA ACT CTG AAC ACA	672
Leu Tyr Asn Phe Ser Asn Thr Gly Asn Pro Asp Pro Thr Leu Asn Thr	
210 215 220	
ACA TAC TTA GAA GTA TTG CGT GCA AGA TGC CCC CAG AAT GCA ACT GGG	720
Thr Tyr Leu Glu Val Leu Arg Ala Arg Cys Pro Gln Asn Ala Thr Gly	
225 230 235 240	
GAT AAC CTC ACC AAT TTG GAC CTG AGC ACA CCT GAT CAA TTT GAC AAC	768
Asp Asn Leu Thr Asn Leu Asp Leu Ser Thr Pro Asp Gln Phe Asp Asn	
245 250 255	
AGA TAC TAC TCC AAT CTT CTG CAG CTC AAT GGC TTA CTT CAG AGT GAC	816
Arg Tyr Tyr Ser Asn Leu Leu Gln Leu Asn Gly Leu Leu Gln Ser Asp	
260 265 270	
CAA GAA CTT TTC TCC ACT CCT GGT GCT GAT ACC ATT CCC ATT GTC AAT	864
Gln Glu Leu Phe Ser Thr Pro Gly Ala Asp Thr Ile Pro Ile Val Asn	
275 280 285	
AGC TTC AGC AGT AAC CAG AAT ACT TTC TTT TCC AAC TTT AGA GTT TCA	912
Ser Phe Ser Ser Asn Gln Asn Thr Phe Phe Ser Asn Phe Arg Val Ser	
290 295 300	



(ix) FEATURE:

(A) NAME/KEY: promoter

(B) LOCATION:1..1532

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide

(B) LOCATION:1533..1609

(ix) FEATURE:

(A) NAME/KEY: exon

(B) LOCATION:1533..1751

(ix) FEATURE:

(A) NAME/KEY: exon

(B) LOCATION:2383..2574

(ix) FEATURE:

(A) NAME/KEY: exon

(B) LOCATION:3605..3769

(ix) FEATURE:

(A) NAME/KEY: exon

(B) LOCATION:4033..4516

(ix) FEATURE:

(A) NAME/KEY: intron

(B) LOCATION:1752..1782

(ix) FEATURE:

(A) NAME/KEY: intron

(B) LOCATION:2575..3604

(ix) FEATURE:

(A) NAME/KEY: intron

(B) LOCATION:3770..4032

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION:1533..1751

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION:2383..2574

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION:3605..3769

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION:4033..4516

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

TAGATAAAAA AATGGGATAT AATTTTTCTC AGATGTTGTT TATACTGTTT TTTTAATCAG	60
AATTAAAATT CCTCTTTAAT TATCGACATA ATTTTTTTTG GTGAATATTA TCGACATAAT	120
TATTTAATAC AAATTTTTAT TGTACATAGA AGTGATACTT CAATTTTAAT ATTGGAGAAC	180

AGTACGAAAA CATAAAAAAA CTGTTATTAG AAGAAAAAAA TATATGGAAA AGGTTAGCTA	240
CATATATTAG CTAAATTAGT TGTCTAATT GGCTATATAA ACCCTATTGT ACTCTTTGTA	300
ATCTCACCTT TTTCATTTAA ATACATTCT ACTTTTAAAG TTCTATATTT TCTCTCAATT	360
TTCTTCGATA AACCATGAAA TTAAACATGG TATATCAGCG ATACCACCCA CTTTGAAAGC	420
CATGTATGGC TAGTATGGGC AGCCAAAATT TGCCCTGGTT CAAGCAAAGC AAGTGTTTAT	480
ATAGATGTGA CTTTTGTTGA GGAATCATG CCAATGGTAC TGATTGTGAA ACTGAGAAAA	540
CTAATTTGGA GAATTTGAAT TATGATCATT AAATACTCCT CTCCTGACTA CCTTCGTCCC	600
TCAAATTTGT ACCATCATT TTTCCCAAAA ATTTGATTAC AATGCACTAA TTAATGAATG	660
TTTCTTACAT TATCATATTA TCATATCTGA CATTTTGTTT TTAATTTTAA TAATAATTAT	720
TTTAAAAAGT CATAATGCA AATAATTTTT TAATAGTTTA CAGTTAAATT TTTACAGTAA	780
AAATGCATGA AAATTAACT TTATTTTCC AAGTCATCAT TTAGTCAAAT CCCAAAACAA	840
TGATTATTTT TTGCAAATGA ATGTTTATTG AACATTTAAA TGTAGCCTAA TTAATTCTGG	900
TTATGGTGTC AATGTTCCAA AACCTAATGC AAGATCTTAG CAAGTACATA CATAGATCTA	960
ATTTTAACT TATCTTTACG CAAGAGATAT AAAGATTATA CATCTAGTTT TAAACATTAA	1020
CTTTTGTTTT TGTGTTAAAA AACAGTAACA TTTCTTAAT TTTGTAGAGT GACGTGCTCC	1080
AACCATATTA ACGAAGATTT TAATTGGTAT TCAAGTTCAT GAACTTAGTA AATAAGTTTT	1140



GGTCTTCAGT TTTCAATTTT CATTACAACA TTTATGTAAA ATATCAACGT TTTCTGAAAT 1200

TTGTTGCTTG TGTGCTCCAA CCACATTTAA GAGATTATAG AAATTAATTT TCAAGAAGAT 1260

AATGATTCCCT ACTCTTGCTG GCCCTACCAT AGTACAATAA ATCCACTCAT AAATCAACAA 1320

GTCGTCGTCA TAGGCAATTG GGCATCATAT CATAACAAT ACGTACGTGA TATTATCTAG 1380

TGTCTCTCAG TTTACTTTAT GAGAAATTAT TTTTCTTTAA AAAAAGTTAA TTAATAAAAA 1440

CATTTGCGAT ACCGTGAGTT ACAAGAAATC CGCCGAATTC ATCTCTATAA ATAAAAGGAT 1500

CTATATGAGA GGTAATCA TATTAECTCA AA ATG GGT TCC ATG CGT CTA TTA 1553

Met Gly Ser Met Arg Leu Leu

355

GTA GTG GCA TTG TTG TGT GCA TTT GCT ATG CAT GCA GGT TTT TCA GTC 1601

Val Val Ala Leu Leu Cys Ala Phe Ala Met His Ala Gly Phe Ser Val

360 365 370 375

TCT TAT GCT CAG CTT ACT CCT ACG TTC TAC AGA GAA ACA TGT CCA AAT 1649

Ser Tyr Ala Gln Leu Thr Pro Thr Phe Tyr Arg Glu Thr Cys Pro Asn

380 385 390

CTG TTC CCT ATT GTG TTT GGA GTA ATC TTC GAT GCT TCT TTC ACC GAT 1697

Leu Phe Pro Ile Val Phe Gly Val Ile Phe Asp Ala Ser Phe Thr Asp

395 400 405

CCC CGA ATC GGG GCC AGT CTC ATG AGG CTT CAT TTT CAT GAT TGC TTT 1745

Pro Arg Ile Gly Ala Ser Leu Met Arg Leu His Phe His Asp Cys Phe

410 415 420

GTT CAA GTACGTACTT TTTTTTTTCC TTCCAAAATG CCCTGCATAT TTAACAAGAT 1801

Val Gln

425

TGCTTTGTTC ACCTAGAAAA ATGTGTTTTT TTCAACGATC TTACGTACGT TTGTTTGTT 1861

TGAAAAATAA ATCAGAAAGA GATCAAGAAA ATAGCTAGAA AGAAAGCAAC GTTTTTTTAA 1921

AAGGTATTTA GTGTGAGAAA AATATTAAAA CTGAAGAGAA AGAAATTAAA TAAGCTTTTC 1981

TTGAATGATA TTTACATGTC TTATTAACCT AAAGTCACCT TTTTCTTTA AGTTGTGCTT 2041

GAAGAAAAAA GATGTCTTTC AGTTTAGTTT TGATTAATGC TAATTATATT TTTAATTAAT 2101

TAATTAATAC TATATATCTA TTTACCATAT TAATTATTAC TATATTTTCAAT GATGACAACA 2161

GACAAGTATT CTAAAGAGGT ATCGGTAGAT GATTAATTTT TTTATAAAAA AATCTTTTGC 2221

GTGTATAGAT ATTCTTTTAT AATTGGTGCA GAAACTTGTA ATGCTAATTG CAATTAATCT 2281

TACATTGATT AACTAATAGC TATAATCAAT ATTTAGGTTA GGTATAGGAG ACAAATCAAG 2341

TGATCTGAAC AAATTAAGTT GTTATATTTG CATTGTGACA G GGT TGT GAT GGA 2394

Gly Cys Asp Gly

1

TCA GTT TTG CTG AAC AAC ACT GAT ACA ATA GAA AGC GAG CAA GAT GCA 2442

Ser Val Leu Leu Asn Asn Thr Asp Thr Ile Glu Ser Glu Gln Asp Ala

5

10

15

20

CTT CCA AAT ATC AAC TCA ATA AGA GGA TTG GAC GTT GTC AAT GAC ATC 2490

Leu Pro Asn Ile Asn Ser Ile Arg Gly Leu Asp Val Val Asn Asp Ile

25

30

35

AAG ACA GCG GTG GAA AAT AGT TGT CCA GAC ACA GTT TCT TGT GCT GAT 2538

Lys Thr Ala Val Glu Asn Ser Cys Pro Asp Thr Val Ser Cys Ala Asp

40

45

50

ATT CTT GCT ATT GCA GCT GAA ATA GCT TCT GTT CTG GTAATTAATA 2584

Ile Leu Ala Ile Ala Ala Glu Ile Ala Ser Val Leu

55

60

ACTCCTAATT AATTCCCAAC CATTAAAAAG TTGCATGATT GGATTCAAAA TTCTATGGTA 2644

TTGGGGTTCT GATATAAATT TGTAATTAAA TTGCACTAAA AAAAATTATC ATATACTTTT 2704

AATAAAAAAA ATTTATCTAA TTTAATTTAT TATTAAACT ATTTTAAAA TTCAATCCTA 2764

ACTCTTTTTT AATCGGAGCA TGTAAGCTGG CACCCACCGT ATATCGTTGG AAGATGCTAT 2824

AAAACCATTT AATTAATGGA TGGAATCAGT CAAAACATTT AATTCAAAAT ACTCTTAATT 2884

GTGATTAGTA ATCATGTTCTG GGCAAGTTAC GTTGTGTATA ATTAATTGA CTTAATCAGA 2944

TAAAAAACA AATGGACGCA AGCCGGTTGG TATAGATATC ACTGGCCTGT AGAATATGTG 3004

GTTTTTCACG TTAAATAAA AGCTAGCTAC TATATTATAT TTAGTCTTTT TTTTCTTAA 3064

ACCCATTTAA CGTGATTTAT TGA CTGTGAA ACATGTTTCC ACACACAGGC TTAGAACTC 3124

CTCGCAACTA ACATCTCCAA AATTTGACTA TTTATTTATG AAGATAATTC ATCTATGATG 3184

TTCAACTCTA TTATATATAT GTATCATCGC AGTATTAAGA ATTATAATAG TCAAATATAG	3244
AAGTATATCG GGTAATGTA GTTGCATGTG CGACCTGTTT CGTGTAAT GCTTATTCTA	3304
TATAGCTTTT TTTATTGGAA AATAACGATG AACTAAAAAC GAAAGGGTAT CATATAGTTT	3364
GACTTTTATG TTAGAGAGAG ACATCTTAAT TTGGTCATAT GTTAAATAAT TAATTACAAT	3424
GCATACACAA ATATTTATGC CATATCTAAA AAATGATAAA ATATCATAGG TATACTCAAC	3484
TATATGATAT CCCCATACA GAAATTGTAC TTTCTTCAG GCAATGAACT TAACATTTCT	3544
GTTTGCTAAA AACAAACATC CACTTAAAGT GGTTCACAT ATTTATGTAA TAATTTACAG	3604
GGA GGA GGT CCA GGA TGG CCA GTT CCA TTA GGA AGA AGG GAC AGC TTA	3652
Gly Gly Gly Pro Gly Trp Pro Val Pro Leu Gly Arg Arg Asp Ser Leu	
1 5 10 15	
ACA GCA AAC CGA ACC CTT GCA AAT CAA AAC CTT CCA GCA CCT TTC TTC	3700
Thr Ala Asn Arg Thr Leu Ala Asn Gln Asn Leu Pro Ala Pro Phe Phe	
20 25 30	
AAC CTC ACT CAA CTT AAA GCT TCC TTT GCT GTT CAA GGT CTC AAC ACC	3748
Asn Leu Thr Gln Leu Lys Ala Ser Phe Ala Val Gln Gly Leu Asn Thr	
35 40 45	
CTT GAT TTA GTT ACA CTC TCA GGTATACATA ATCAATTTTT TATTTGCTAT	3799
Leu Asp Leu Val Thr Leu Ser	
50 55	
TAGCTAGCAA TAAAAAGTCT CTGATACAGA CATATTTAGA TAAATTAATT TCTCCATAAA	3859

CATTTATAAT AAAATTATCA ATTTATGTAC TTAAAAATTA TGGATTGAAG CTCTTTTCAT 3919

CCAACTTTTA CTAAAGTTAA GGTGCATATA ATATAAATA AACTATCTCT TGTTTCTTAT 3979

AAAAAGATTG AAGATAAGTT AAAGTCTACT TATAAATCAT TAATATATGT ATA GGT 4035  
Gly  
1

GGT CAT ACG TTT GGA AGA GCT CGG TGC AGT ACA TTC ATA AAC CGA TTA 4083  
Gly His Thr Phe Gly Arg Ala Arg Cys Ser Thr Phe Ile Asn Arg Leu  
5 10 15

TAC AAC TTC AGC AAC ACT GGA AAC CCT GAT CCA ACT CTG AAC ACA ACA 4131  
Tyr Asn Phe Ser Asn Thr Gly Asn Pro Asp Pro Thr Leu Asn Thr Thr  
20 25 30

TAC TTA GAA GTA TTG CGT GCA AGA TGC CCC CAG AAT GCA ACT GGG GAT 4179  
Tyr Leu Glu Val Leu Arg Ala Arg Cys Pro Gln Asn Ala Thr Gly Asp  
35 40 45

AAC CTC ACC AAT TTG GAC CTG AGC ACA CCT GAT CAA TTT GAC AAC AGA 4227  
Asn Leu Thr Asn Leu Asp Leu Ser Thr Pro Asp Gln Phe Asp Asn Arg  
50 55 60 65

TAC TAC TCC AAT CTT CTG CAG CTC AAT GGC TTA CTT CAG AGT GAC CAA 4275  
Tyr Tyr Ser Asn Leu Leu Gln Leu Asn Gly Leu Leu Gln Ser Asp Gln  
70 75 80

GAA CTT TTC TCC ACT CCT GGT GCT GAT ACC ATT CCC ATT GTC AAT AGC 4323  
Glu Leu Phe Ser Thr Pro Gly Ala Asp Thr Ile Pro Ile Val Asn Ser  
85 90 95

TTC AGC AGT AAC CAG AAT ACT TTC TTT TCC AAC TTT AGA GTT TCA ATG 4371  
Phe Ser Ser Asn Gln Asn Thr Phe Phe Ser Asn Phe Arg Val Ser Met  
100 105 110

ATA AAA ATG GGT AAT ATT GGA GTG CTG ACT GGG GAT GAA GGA GAA ATT 4419  
Ile Lys Met Gly Asn Ile Gly Val Leu Thr Gly Asp Glu Gly Glu Ile  
115 120 125

CGC TTG CAA TGT AAT TTT GTG AAT GGA GAC TCG TTT GGA TTA GCT AGT 4467  
Arg Leu Gln Cys Asn Phe Val Asn Gly Asp Ser Phe Gly Leu Ala Ser  
130 135 140 145

GTG GCG TCC AAA GAT GCT AAA CAA AAG CTT GTT GCT CAA TCT AAA TAA 4515  
Val Ala Ser Lys Asp Ala Lys Gln Lys Leu Val Ala Gln Ser Lys \*  
150 155 160

ACCAATAATT AATGGGGATG TGCATGCTAG CTAGCATGTA AAGGCAAATT AGGTTGTAAA 4575

CCTCTTTGCT AGCTATATTG AAATAAACCA AAGGAGTAGT GTGCATGTCA ATTCGATTTT 4635

GCCATGTACC TCTTGGAATA TTATGTAATA ATTATTTGAA TCTCTTTAAG GTA CTTAATT 4695

AATCA 4700